Metabarcoding and Metagenomics 2: 1–17
DOI 10.3897/mbmg.2.22467

Research Article

Dropping Hints: Estimating the diets of livestock in rangelands using DNA metabarcoding of faeces

Timothy R. C. Lee[1], Yohannes Alemseged[2], Andrew Mitchell[1]

1 *Australian Museum, Sydney, Australia.*

2 *NSW Department of Primary Industries, Trangie, Australia.*

Corresponding author: *Timothy R. C. Lee* (timothy.lee@austmus.gov.au)

Academic editor: *Emre Keskin* | **Received** 23 November 2017 | **Accepted** 17 February 2018 | **Published** 14 March 2018

Abstract

The introduction of domesticated animals into new environments can lead to considerable ecological disruption, and it can be difficult to predict their impact on the new ecosystem. In this study, we use faecal metabarcoding to characterize the diets of three ruminant taxa in the rangelands of south-western New South Wales, Australia. Our study organisms included goats (*Capra aegagrus hircus*) and two breeds of sheep (*Ovis aries*): Merinos, which have been present in Australia for over two hundred years, and Dorpers, which were introduced in the 1990s. We used High-Throughput Sequencing methods to sequence the *rbcL* and ITS2 genes of plants in the faecal samples, and identified the samples using the GenBank and BOLD online databases, as well as a reference collection of sequences from plants collected in the study area. We found that the diets of all three taxa were dominated by the family Malvaceae, and that the Dorper diet was more diverse than the Merino diet at both the family and the species level. We conclude that Dorpers, like Merinos, are potentially a threat to some vulnerable species in the rangelands of New South Wales.

Keywords

Metabarcoding, *rbcL*, ITS2, *Ovis aries*, diet metabarcoding, barcode references

Introduction

Estimating the diet of an animal by analysing its faeces has many advantages: it is non- invasive and does not require time consuming observation of the animal. Microscopic analysis of plant fragments in animal faeces has long been used to investigate diets (Todd and Hansen 1973) but DNA-based dietary analysis is increasingly utilised in a broad range of species (Höss et al. 1992, Pompanon et al. 2011). These new methods have been found to recover greater diversity in faecal samples than traditional ones (Soininen et al. 2009). Most studies of diet from faecal material have focused on carnivores (Jarman et al. 2013, Shehzad et al. 2012), however, several studies have shown that the technique can work in herbivores (Soininen et al. 2009, Hibert et al. 2013, Symondson and Harwood 2014). In this study, we use faecal DNA barcoding analysis to characterize the diets of economically and ecologically significant Dorper and Merino sheep (*Ovis aries*) and the goat (*Capra aegagrus hircus*).

In order to determine an animal's diet using DNA recovered from its faeces, there must exist a reference database of sequences recovered from potential dietary species, against which sequences recovered from faeces can be compared. DNA barcoding (Hebert et al. 2003) is starting to address this requirement, by building reference databases for all species through the sequencing of standardized genomic regions, while adhering to rigorous standards for collateral data and vouchering of the sequenced specimens in museum/herbarium collections (Mitchell 2008). While the mitochondrial COI gene is a very effective DNA barcode standard for animals (Coissac et al. 2016), there is no single candidate marker in plants with the same characteristics. The DNA barcode standard for plants therefore comprises two chloroplast gene fragments, the ribulose-1,5-bisphosphate carboxylase/oxygenase large-chain (*rbcL*) and maturase K (*matK*, CBOL Plant Working Group et al. 2009). *RbcL* evolves slowly and has been widely used to reconstruct deep evolutionary relationships in plants (Chase et al. 1993).

Copyright *Timothy Lee et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

MatK evolves more rapidly and can distinguish about 70% of plant species (Parmentier et al. 2013). However, *matK* may lack the conserved sites suitable for PCR primer binding (but see Heckenhauer et al. 2016), limiting its utility for barcoding, and currently there are fewer *matK* than *rbcL* sequences available in international databases (~175,000 plant *rbcL* sequences on GenBank against ~150,000 plant *matK* sequences at the time of writing). Furthermore, chloroplast and mitochondrial DNA is inherited only through the maternal lineage, limiting their utility in detecting hybridization and introgression events, all too common in plants.

To overcome these issues, the ITS2 region of the nuclear ribosomal RNA genes was proposed as a supplementary DNA barcode marker for plants (Chen et al. 2010, Yao et al. 2010). ITS2 has the advantages that the PCR primers are "universal," and the length of the region is only ~230 bp in land plants, with ~95% of the variation between 100 bp and 350 bp (Yao et al. 2010). This means it is usually short enough to allow its PCR amplification from samples with fragmented DNA. Additionally, as part of the nuclear genome, ITS2 evolves independently of *rbcL* and *matK*.

Once a database of standard marker sequences from potential dietary taxa has been assembled, the same marker must be recovered and sequenced from faecal samples, where DNA is fragmented and species-mixed. This requirement is being met by high- throughput sequencing (HTS) technologies, which have proliferated since the turn of the century. Compared to the older Sanger sequencing method (Sanger et al. 1977), HTS technologies focus on sequencing shorter DNA fragments, which are joined together by computer analysis to form longer sequences. This enables fragmented template DNA to be sequenced, and different individual organisms' sequences in a sample can be determined simultaneously. There are a number of HTS technologies currently competing for market share, each with their own strengths and weaknesses (Glenn 2011, Goodwin et al. 2016, Levy and Myers 2016).

Combining these two technology-driven fields, DNA barcoding and HTS, gives us DNA metabarcoding (Taberlet et al. 2012), the ability to construct a species list from a sample of environmental DNA. This is well suited to the task of reconstructing sheep diets from the plant DNA present in their faeces. As of 2014, Australia is the world's second largest producer of sheep, after China, although the size of the Australian flock has declined in the last 10 years (Food and Agriculture Organization of the United Nations, 2017). Merino sheep, having originated in Spain, have been present in Australia since the late 18th century and comprise about 75% of Australia's sheep population (Royal Agricultural Society of Tasmania, 2016). A more recent addition to the Australian flock is the Dorper, which was developed in South Africa in the 1930s by crossing the ***Dor***set Horn and Blackheaded ***Per***sian sheep. Dorpers were first brought to Australia in 1996 (Dorper Sheep Society of Australia, 2017), where it was thought their drought-tolerant physiology would be well suited. While the less selective diet of Dorpers relative to Merinos has been previously noted (reviewed in Brand 2000), no faeces metabarcoding study has yet been conducted to investigate this. Additionally, while research has been undertaken on the management of Dorpers in Australia (Alemseged and Hacker 2014), the potential implications of the broad Dorper diet on endangered flora has not yet been investigated.

This project was designed to test the utility and practicality of DNA metabarcoding for dietary analysis of different livestock breeds. We focused on the Merino and Dorper, as well as goats as a control, in south-western New South Wales (NSW).

Our aims were:

1. Use a DNA metabarcoding approach to compare the diets of three ruminants: Merino sheep (Australia's most popular breed, introduced over 200 years ago); Dorper sheep (a recently introduced breed with a less well characterised diet in Australia) and goats (a control) in the rangelands of south-western NSW.
2. To determine whether Dorpers are a potential threat to vulnerable flora.

Material and methods

Study Area

Dorper, Merino and goat faeces samples were collected from 16 different sheep or goat rearing properties in south-western New South Wales, Australia (Suppl. material 1). To minimise environmental variation between Dorper and Merino sites, they were selected in pairs, with the Dorper and Merino sites in each pair being no further than 50km from one another (with the exception of the Aston site, which was an extra unpaired Merino site). Faeces were collected wet within 30mins of being deposited using flame sterilised tweezers, and stored in ethanol for transport. The sampling sites are in the Murray Darling Depression and Darling Riverine Plains bioregions, as determined by the Interim Biogeographical Regionalisation for Australia 7 (IBRA7). These regions have a semiarid climate with predominately winter rainfall.

The reference DNA barcode library and Sanger sequencing methods

We opted to focus on the ITS2 and *rbcL* markers in this study. *RbcL* was included because it is a standard plant barcoding gene, and there already exists a large online database of sequences. ITS2 was included as it evolves more quickly than *rbcL*, increasing its variability, and because it has a different mode of inheritance (nuclear rather than chloroplast). We chose not to focus on *matK* due to its smaller online sequence database, and also because its high degree of variability in primer binding sites can decrease the likelihood of successful primer

binding, particularly in species-mixed samples with high taxonomic diversity (Heckenhauer et al. 2016, Hollingsworth et al. 2011). To identify the *rbcL* and ITS2 sequences from our samples, we used two online sequence databases: the Barcode of Life Data Systems (BOLD, Ratnasingham and Hebert 2007) and NCBI's GenBank (Benson et al. 2012). BOLD was launched in 2005 as an online sequence database specifically dedicated to DNA barcoding applications, and it is focused on collecting sequences that are BARCODE compliant, meaning they are of high quality and come from well curated voucher specimens (Hanner 2012). The GenBank and BOLD databases currently have only a small number of sequences from plants found in the study area. We therefore supplemented the BOLD database with sequences from 24 plant reference samples collected from the study sites (Table 1). These species were chosen as they were the most common plants in the study site that were likely targets for browsers. The complete species diversity of the site, however, is unknown, and there may be many species from the study area that remain unbarcoded.

Of the 24 reference samples collected, five could only be identified to genus. These five genera contain species that occur in the study area and are listed in NSW as threatened on at least one of two different lists; the Australian Environment Protection and Biodiversity Conservation Act 1999 national threatened flora list (Department of the Environment and Energy Commonwealth of Australia, 2018b), and the Buloke Woodlands of the Riverina and Murray-Darling Depression Bioregions list (Department of the Environment and Energy Commonwealth of Australia, 2018a), which is specific to the study area. These species are *Sclerolaena napiformis* (Endangered), *Austrostipa metatoris* (Vulnerable), *A. nullanulla* (Endangered) and *A. wakoolica* (Endangered) on the national list, and *Ptilotus erubescens* (Rare/Threatened), *P. exaltatus* var. *semilanatus* (Endangered), *Austrostipa exilis* (Rare), *A. gibbosa* (Rare), *A. puberula* (Rare), and *Sclerolaena napiformis* (Threatened/Endangered) on the Buloke Woodlands list. This makes it possible that some of the specimens included in the reference collection are threatened. None of the reference specimens identified to species level are threatened in NSW.

Table 1. Plant samples used for DNA barcode library construction. Botanical name is the initial identification before DNA sequencing. See DNA barcode results in Suppl. material 2 and Suppl. material 3.

ID	Common name	Botanical name	Date collected	Site	Lat (-)	Long
T01	Streaked Poverty-Bush	*Sclerolaena tricuspis*	26/07/2011	K-Tank	31.86261	141.83269
T02	Manna Wattle	*Acacia microcarpa*	27/07/2011	C-lake	33.08461	143.47333
T03	Black Bluebush	*Maireana pyramidata*	25/05/2011	Mazar	32.76723	141.06483
T04	Quena	*Solanum esuriale*	25/05/2011	Coombah (goat)	33.03057	141.73007
T05	Rosewood	*Heterodendrum oleifolium*	27/07/2011	C-lake	33.09926	143.55813
T06	Erect Mallee Bluebush	*Maireana pentatropis*	27/07/2011	C-lake	33.09926	143.55813
T07	Turpentine	*Eremophila sturtii*	27/07/2011	C-lake	33.08461	143.47333
T08	Pearl Bluebush	*Maireana sedifolia*	28/07/2011	near Ivanhoe	33.29779	143.928669
T09	Common Bottlewashers	*Enneapogon avenaceus*	26/07/2011	K-tank	31.86261	141.83269
T10	Copper burr	*Sclerolaena* sp.	28/07/2011	Baymore	33.46041	143.17698
T11	Bladder Saltbush	*Atriplex vesicaria*	28/07/2011	Baymore	33.46041	143.17698
T12	Prickly Wattle	*Acacia victoriae*	25/05/2011	Mazar	32.76723	141.06483
T13	Box Grass	*Paspalidium constrictum*	25/05/2011	Coombah	33.03057	141.73007
T14	Harlequin Mistletoe	*Lysiana exocarpi*	25/05/2011	Coombah	33.03057	141.73007
T15	Twiggy Sida	*Sida intricata*	13/07/2011	Warrananga	33.70348	141.73955
T16	Kerosene Grass	*Aristida contorta*	12/07/2011	Bunnerungee	33.54223	141.73538
T17	Ruby Saltbush	*Enchylaena tomentosa*	12/07/2011	Aston Merino	33.2956	142.326488
T18	Ward's Weed	*Carrichtera annua*	7/06/2011	Orana	32.76421	144.04387
T19	*Austrostipa* sp.	*Austrostipa* sp.	7/06/2011	Eurella	32.63605	144.24124
T20	Cannon-Ball	*Sclerolaena paradoxa*	7/06/2011	Orana	32.76421	144.04387
T21	Crows foot	*Erodium* sp.	27/06/2012	Eurella	32.6	144.27
T22	Medic	*Medicago* sp.	27/06/2012	Eurella	32.6	144.27
T23	Ptilotus	*Ptilotus* sp.	29/05/2012	Kimberley	32.85	141.15
T24	Pop saltbush	*Atriplex holocarpa*	28/06/2012	Baymore	33.44	143.15

Although we did not attempt to sequence *matK* in the faeces samples, the *matK* sequences of reference plant samples were included in this study because they could facilitate species identification at a future date. Published PCR primers for *rbcL*, *matK* and ITS2 (Table 2), yielding amplicons of approximately 600, 800 and 460 bp respectively, were used for PCR amplification and sequencing of reference plant samples.

For amplifying *rbcL* and *matK*, each 15 µL PCR contained: 2 µL of DNA, 0.025 µL MilliQ water, 1.5 µL of 10x reaction buffer at 10 mM, 1.5 µL of $MgCl_2$ at 25 mM, 0.3 µL of dNTPs at 10 mM, 0.075 µL of Platinum Taq at 5 U/µL and 0.3 µL each of forward and reverse primer at 5 µM. Cycling conditions for *rbcL* and *matK* were as follows: 94°C for 2 min; 94°C for 30 s, 55°C for 30 s and 72°C for 1 min x 5 cycles; 94°C for 30 s, 54°C for 30 s and 72°C for 1 min x 30 cycles; and 72°C for 7 mins. The ITS2 amplification followed the same protocol, except that 35 PCR cycles were performed, and the annealing temperature was 55°C. Sanger sequencing was carried out by Macrogen Inc. (Seoul, South Korea) using a high throughput Applied Biosystems 3730XL sequencer. Peak calling and contig assembly was carried out in Geneious 10.0.2 (Kearse et al. 2012).

Table 2. PCR primers used. Notes: 1 = used for DNA barcode library data collection, 2 = used for high- throughput sequencing of faecal pellets, 3 = used in combination with *rbcL*ajf634r

Gene	Primer name	Amplicon size (excl. primers)	Note	Sequence (5'-3')	Source
matK	1RXkim	~800 bp	1	ACCCAGTCCATCTGGAAATCTTGGTNC	Ki-Joong Kim (unpublished)
matK	3FXkim		1	CGTACAGTACTTTTGTGTTTACGNG	Ki-Joong Kim (unpublished)
rbcL	*rbcL*af	607 bp	1	ATGTCACCACAAACAGAGACTAAAGC	Fazekas et al. (2008)
rbcL	*rbcL*ajf634r		1, 2	GAAACGGTCTCTCCAACGCAT	Fazekas et al. (2008)
rbcL	*rbcL*-AM2f	247 bp	2, 3	AAYGTYTTTGGKTTCAARGC	This study
ITS2	S2F	~460 bp	1, 2	ATGCGATACTTGGTGTGAAT	Chen et al. (2010)
ITS2	S3R		1, 2	GCTTCTCCAGACTACAAT	Chen et al. (2010)

We estimated phylogenetic trees for all three genes using the sequences recovered from the reference samples, in order to explore the diversity of the three markers among our reference sequences and to check for any sequence clusters from different taxa with low divergence. All three trees were rooted with a single outgroup sequence from the Scots Pine (*Pinus sylvestris*), downloaded from GenBank. Phylogenetic analysis was carried out in Geneious 10.0.2 using the MrBayes v3.2.6 plugin (Ronquist et al. 2012). DNA substitution models for the *matK*, *rbcL* and ITS2 genes were estimated in ModelGenerator v0.84 (Keane et al. 2006), using the model with the minimum negative log likelihood under the Akaike Information Criterion 1. Posterior distributions of parameters were estimated using Markov Chain Monte Carlo sampling. All three analyses were carried out with a chain length of 1,100,000 with samples drawn every 200 steps. The first 100,000 steps were discarded as burn-in. Convergence was tested for by verifying that the standard deviation of split frequencies had dropped below 0.01 in each case.

High-Throughput Sequencing (HTS)

Twenty-four faecal samples were chosen to include an approximately equal number of the three treatments (Merino, Dorper or goat) from across the study area (Table 3). Two pairs of faeces samples came from the same sheep: sample 14 and 19 were from the same Dorper, and sample 15 and 20 were from the same Merino. These duplicate samples served as internal controls for among-sample variation in DNA sequence diversity. Overall, the 24 samples comprised eight Merino plus one repeat, seven Dorpers plus one repeat and seven goat samples. DNA extractions were performed at the Australian Museum using both a Bioline Isolate II Plant DNA Isolation kit and a Qiagen DNeasy (Animal) Tissue Kit.

All DNA extractions were quantitated using a Nanodrop spectrophotometer. The Qiagen kit yielded DNA extractions at higher DNA quantity and purity, so these DNA extractions were sent to the Ramaciotti Centre for Gene Functional Analysis (University of New South Wales, Sydney) for further processing.

Published plant DNA metabarcoding studies either have sampled only fresh plant material to generate large PCR amplicons for DNA sequencing, or have utilised alternative chloroplast genes, such as *trnL*, for which tried and tested primers yielding small fragments were available (Valentini et al. 2009). Short amplicons are favoured in faecal metabarcoding studies as DNA of dietary organisms is fragmented during digestion, and short markers are more likely to be successfully amplified in the fragmented DNA (King et al. 2008). No published primers could be found that targeted amplicons of the appropriate sizes (<300 bp for faecal samples, King et al. 2008) for the plant DNA barcode standard genes, *matK* and *rbcL*. Therefore a new PCR primer (*rbcL*-AM2f) was designed to amplify a 247 bp fragment of *rbcL*, when used in combination with one of the standard barcode

Table 3. Sheep faecal pellet samples used. Locations of properties can be found in Suppl. material 1.

Sample ID	Property (Paddock if different to stock)	Stock (Notes)
1	Warrananga	Merino
2	Orana	Goat
3	Mazar (Merino)	Goat (wet + dry)
4	Mazar (Merino)	Merino (wet + dry)
5	K-Tank	Goat
6	Eurella	Dorper
7	C-Lake	Merino
8	Kimberley	Dorper
9	Kimberley	Goat
10	Coombah	Goat
11	Popiltah	Dorper
12	Coombah	Merino
13	C-Lake	Goat
14	Bunnerungee	Dorper
15	Bunnerungee	Merino
16	Clevedale	Dorper
17	Baymore	Dorper
18	Bonton	Merino
19	Bunnerungee	Dorper (repeat of 14)
20	Bunnerungee	Merino (repeat of 15)
21	Avoca	Dorper
22	Aston	Goat
23	Aston	Merino
24	Avondale	Merino

primers, *rbcL*ajf634r (Fazekas et al. 2008). *RbcL*-AM2f was designed manually in Geneious, using *rbcL* sequences from the 24 reference samples as a reference. The primer was tested with the *rbcLajf634r* reverse primer on the 24 reference samples and the fragment amplified successfully in all of them. The source, sequences and expected amplicon size of the PCR primers utilised for the HTS part of this study are given in Table 2.

HTS was performed by the Ramaciotti Centre using the Illumina MiSeq platform. The Nextera XT 24-sample preparation kit was used because it facilitates indexing of 24 separate samples, i.e. each sample is labelled with a unique DNA sequence "index" which allows all the sequences from a particular sample to be separated after the sequencing run. The Illumina MiSeq Nano version 2 sequencing kit was used, allowing sequencing of 250 bp from each end of an amplicon. As both amplicons were less than 500 bp long, this allowed for bidirectional sequencing of each template, ensuring that high quality sequence data was obtained.

HTS data analysis

The Illumina high-throughput sequencing output consisted of 1,110,113 forward and reverse reads over the 24 samples, averaging 46,255 reads per sample (number per sample can be found in Suppl. material 4).

Raw Illumina reads were imported into Geneious 10.0.2, filtered for length (>35 bp) and ends of sequences with stretches of Ns or quality lower than 14 were trimmed. In Geneious, bidirectional reads were merged, and any reads that failed to merge were discarded. Reads were then separated into *rbcL* and ITS2 reads by searching for ITS2 primer sequences within the reads in Geneious. The separated reads were then pooled across all samples and screened for Chimeras using the *de novo* uchime2 algorithm implemented in USEARCH v9.2.64 (Edgar et al. 2011).

The non-chimeric reads were assembled into contigs in Geneious. Sequences had primers trimmed with a maximum of 2% gaps and maximum 2 bp gap size. Maximum mismatches per read was 2% for ITS2 and 1% for *rbcL*, due to the higher degree of variation expected in the ITS2 gene (Chen et al. 2010). Following assembly, *rbcL* contigs with a length longer than 300 bp and ITS2 contigs with a length shorter than 300 bp were excluded. Of the ITS2 contigs, 77.1% of those excluded for being under 300 bp were also under 50 bp in length. As they matched well with primer sequences they were likely primer dimers. Five (or where there were fewer than five, all) of the remaining 22.9% of excluded ITS2 contigs from each sheep were BLASTed to determine what they matched to. In all cases, these sequences either had no match, or matched to ribosomal RNA sequences not in-

cluding ITS2, or overlapped less than 100 bp with an ITS2 sequence. We therefore excluded these sequences as none of those tested contained complete ITS2 sequences. ITS2 contigs were screened for 5.8S and 28S sequences, using the 'Annotate from Database' function in Geneious. Annotations were made using an annotated plant sequence downloaded from GenBank, from the species *Canella winterana* (sequence L03844.1) with annotations made at 40% similarity (counting ambiguities as mismatches). Annotated sections were then removed, leaving only the ITS2 sequence. Contigs with a read depth of less than 10 were excluded in both genes in the main analysis, following the 10-minimum read depth protocol of Hibert et al. (2013) and Gebremedhin et al. (2016). We also carried out a supplementary analysis with all the same conditions but at a minimum read depth of 3.

Sequence Identification

Sequence identification using the GenBank database was conducted online, using BLASTN 2.6.0 (Zhang et al. 2000). Identification using BOLD was conducted using the MegaBLAST function in Geneious, searching a downloaded database of all BOLD ITS2 and *rbcL* sequences that was supplemented by our reference database sequences.

Match data from GenBank, consisting of the 100 closest matches for each contig, were then imported into MEGAN6 Community Edition v6.7.0 (Huson et al. 2016) for the assignment of the lowest common ancestor (LCA) of the best matches. The default LCA parameters were used, however, the "minimum percent identity" was changed to 1% in ITS2 and 0.005% in *rbcL*, to account for the relatively higher inter-specific variation in ITS2 (Chen et al. 2010). Match data from BOLD consisted of the 100 closest matches, which was then restricted to matches with 100% query coverage and a minimum percentage identity, 98% for ITS2, 99% for *rbcL*, again to account for the more variable ITS2. Identifications were then assigned for each contig based on the lowest taxonomic level in common, up to family, among all their matching database sequences. This ensured that sequences that could not be identified to species could still be identified to family, but those sequences not identifiable to family were excluded from further analysis.

All sequences identified from each database were then collated into an alignment, one for BOLD identifications and one for GenBank, and neighbour-joining trees produced using the Geneious Tree Builder, with default settings. These trees were used to help determine the identity of OTUs that received no hits in either the BOLD or GenBank databases.

IDs assigned to taxa were compared against the Atlas of Living Australia (ALA) website (Atlas of Living Australia 2017). If the assigned species did not occur in the study zone, or within 500km of it, the ID was changed to '*Genus* sp.', and all such instances are underlined in our results tables. We close to include these matches to taxa that occurred outside the study zone, rather than exclude them entirely as in Sugimoto et al. 2018, as the reference database for our study area is not complete and we felt that matches outside the study area that were congeneric with plants within it were therefore still worth noting. If no member of the genus occurred in the study area, the sequence was only identified to family (this only occurred once, in the BOLD *rbcL* data).

Sheep Diet Analysis

We performed statistical tests on the GenBank sheep diet data to investigate similarities and differences between the breeds. We performed two-tailed t-tests in R v3.3.3 (R Core Team 2017) between the Dorper and Merino samples, based on the family and species diversity in each sample. Data was tested for normality and equality of variance using the Shapiro-Wilk test in R, and Levene's test in the R package 'car' v2.1-6 (Weisberg and Fox 2011). We also performed two ANOSIMs at the family level using the package vegan (Oksanen et al. 2017) in R, one using proportional read data and one using presence/absence data, in both cases using the breed as the independent variable, to test for differences in the proportion of plant families in the diets of the three taxa. We also calculated the Sørensen similarity index for all three pairs of taxa at the species level. Finally, we used the Kruskal-Wallis test implemented in the R to compare the total number of contigs to the number of contigs that could be identified to species in each taxon.

Data resources

The metabarcoding datasets generated and analysed during the current study are available at Figshare: http://doi.org/10.6084/m9.figshare.5309935

The plant reference database generated and analysed during the current study is available at BOLD: http://doi.org/10.5883/DS-SMPRS

Results

DNA barcode reference library

Sequences were obtained for the *rbcL*, *matK* and ITS2 genes for 18, 16 and 15 of the 24 plant reference samples respectively.

Phylogenetic trees were derived from the slower evolving *rbcL* gene (Fig. 1) and the faster evolving *matK* gene (Fig. 2). There were several groups of identical sequences from samples identified morphologically as different species, in *rbcL* (T10 and T17; T03 and T08, and T11 and T24) and *matK* (T01 and T10). *MatK* provided better resolving power than *rbcL*, however *matK* sequences were also unable to distinguish clearly between all samples. A phylogenetic tree was also derived from the nuclear ITS2

gene (Fig. 3). In the ITS2 dataset, no sequences were identical, which might make this marker more useful in identifying sequences to species. In all three trees, specimens identified morphologically as from the same family clustered together with maximum posterior probability in all cases except Fabaceae in *rbcL*, where it was 0.92.

High-throughput sequencing of faecal samples

De novo chimeric sequence detection using the UCHIME algorithm flagged 21.4% of reads as chimeric in ITS2 and 42.1% as chimeric in *rbcL*. Following their removal, the non- chimeric reads were assembled into contigs. These ranged from 66-761 contigs per sheep for ITS2 and 119-1060 contigs per sheep for *rbcL*.

A number of matches on both the GenBank and BOLD databases, particularly in the *rbcL* gene, were for unicellular algae that were likely present in the diet as water contaminants. In all analyses, algal sequences were excluded. In the following, singletons (species or families appearing in only one sample) were also excluded.

Some specimen identifications were changed to "*Genus* sp." based on available information about taxon ranges from ALA. This accounted for the incompleteness of the reference databases, so that matches to species that did not occur within the study zone were instead listed as matching only to genus level. All such cases are clearly marked in Table 4 and Suppl. materials 6, 13, 15. In one case, *Aethionema* sp. in the BOLD *rbcL* data, the genus did not occur in the study area and the match was changed to 'Brassicaceae species'.

GenBank Results

Generally, Dorpers were found to have more diverse diets than Merinos. Based on the GenBank results, at the family level, there were 11 families present in the diets of both breeds, three families found in Merinos not Dorpers, and seven families found in Dorpers not Merinos (Table 5). At the species level, there were 15 species in common between the breeds, six species found in Merinos not Dorpers and 12 species found in Dorpers not Merinos (Table 4). At the family level, the goat diet included 13 families, which was fewer families than in either Merinos or Dorpers. One of these families was unique to goats, two were shared only with Dorpers, and ten families were common to all three taxa. At the species level, the goat diet was also less varied than either sheep, having only 18 species present. Four of these species were unique to goats, two were shared only with Merinos, two were shared only with Dorpers, and 10 species were common to all three taxa (Table 4). Based on the GenBank results, the Sørensen similarity index, at species level between the Merinos and Dorpers was 62.5%. Between Dorpers and goats it was 53.3% and between Merinos and goats it was 61.5%. This indicates that although the Dorper diet is more diverse than Merino diet, those two diets are more similar to one another than either is to the goat diet.

Threatened Species

The list of species found in the Dorper and Merino faeces based on Genbank results was compared to the national list of Australian threatened taxa (Department of the Environment and Energy Commonwealth of Australia, 2018b), and the list of threatened taxa in the Buloke Woodlands region, an endangered habitat in the Riverina and Murray-Darling depression bioregions occurring in and around the study site (Department of the Environment and Energy Commonwealth of Australia, 2018a). Threatened taxa found in the diets of sheep were *Minuria cunninghamii* (Dorper only, listed as 'rare') and *Sida fibulifera* (Dorper, Merino and goat, 'vulnerable') which are threatened in the Buloke Woodlands region. These were species-level matches, but there were also some genus-level matches to threatened species, for sequences identified only to genus level. On the national list, these were *Sclerolaena napiformis* (Merino only, 'endangered'), *Calotis moorei* (Dorper, Merino and goat, 'endangered'), and *Maireana cheelii* (Dorper and Merino, 'vulnerable'). On the Buloke Woodlands list, these were *Sclerolaena napiformis* (Merino only, 'threatened/endangered'), *Maireana cheelii* (Dorper and Merino, 'vulnerable'), *M. excavata* (Dorper and Merino, 'vulnerable') and *M. rohrlachii* (Dorper and Merino, 'rare').

Table 4. Species level taxa (GenBank Data). Names of taxa with asterix were changed based on the distribution of taxa in the study zone.

Merino Only	Dorper Only	Goat Only	Dorper and Merino	Merino and Goat	Dorper and Goat	Dorper, Merino and Goat
Calotis hispidula	*Boerhavia* sp.*	*Chenopodium auricomum*	*Austrostipa nodosa*	*Medicago laciniata*	*Erodium cygnorum*	*Austrostipa nitida*
Cullen australasicum	*Brachyscome ciliaris*	*Haloragis aspersa*	*Lotus* sp.*	*Sclerolaena obliquicuspis*	*Tetragonia* sp.*	*Calotis* sp.*
Sclerolaena sp. 1*	*Erodium cicutarium*	*Haloragis* sp.	*Maireana* sp.*			*Carrichtera annua*
Sclerolaena sp. 2*	*Leiocarpa semicalva*	*Haloragis glauca*	*Salvia* sp.*			*Convolvulus clementii*
	Medicago polymorpha		*Vittadinia sulcata*			*Medicago minima*
	Minuria cunninghamii					*Sclerolaena diacantha*
	Silene sp.*					*Sida fibulifera*
	Sisymbrium erysimoides					*Sida* sp.*
	Sonchus sp.*					*Tetragonia tetragonioides*
	Spergularia tasmanica					*Vittadinia eremaea*

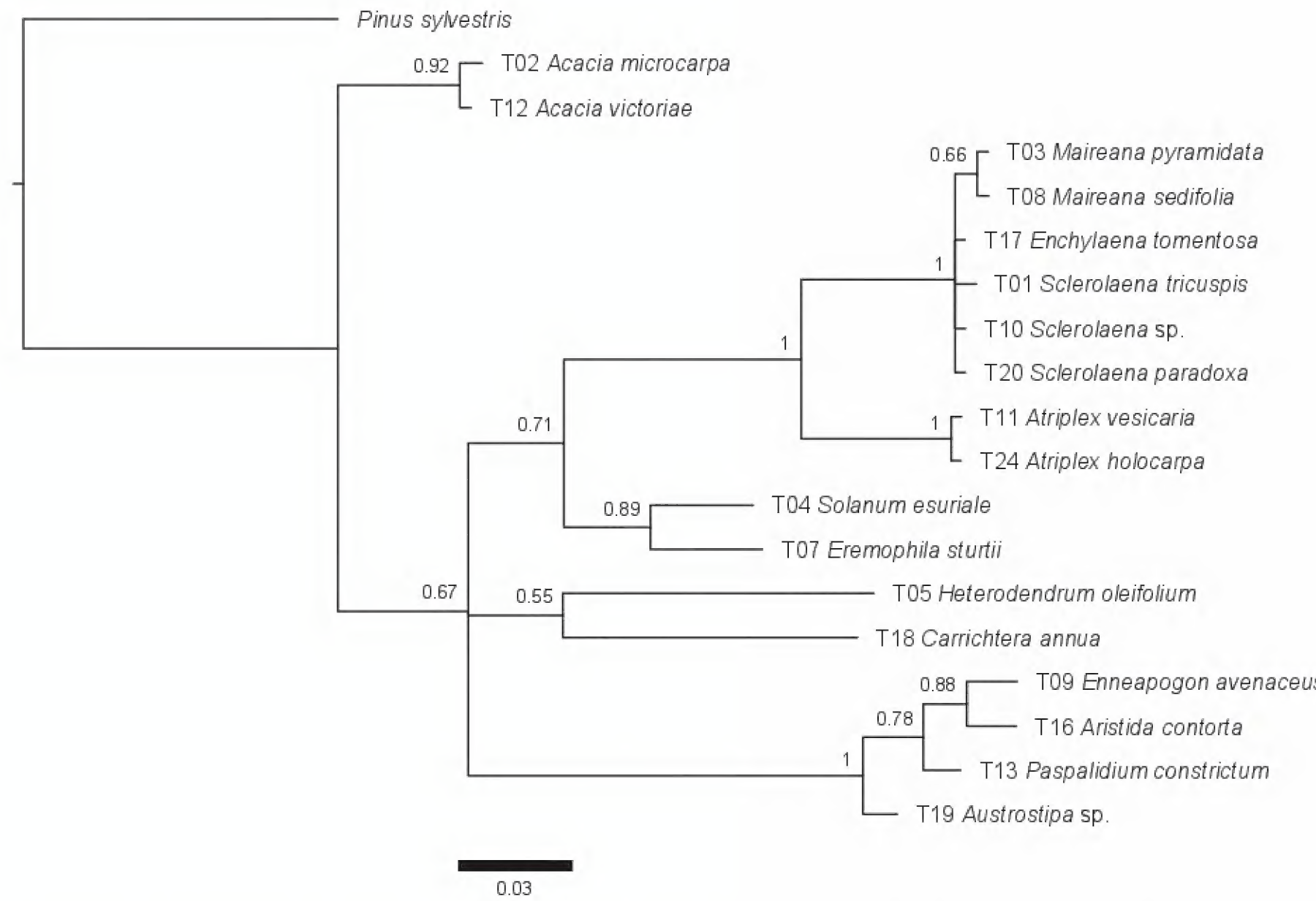

Figure 1. Bayesian tree of *rbcL* gene sequence data. Numbers at nodes are posterior probability values. Scale bar indicates substitutions per site. The tree was rooted with a single *Pinus sylvestris* sequence downloaded from GenBank (accession number AB097775.1).

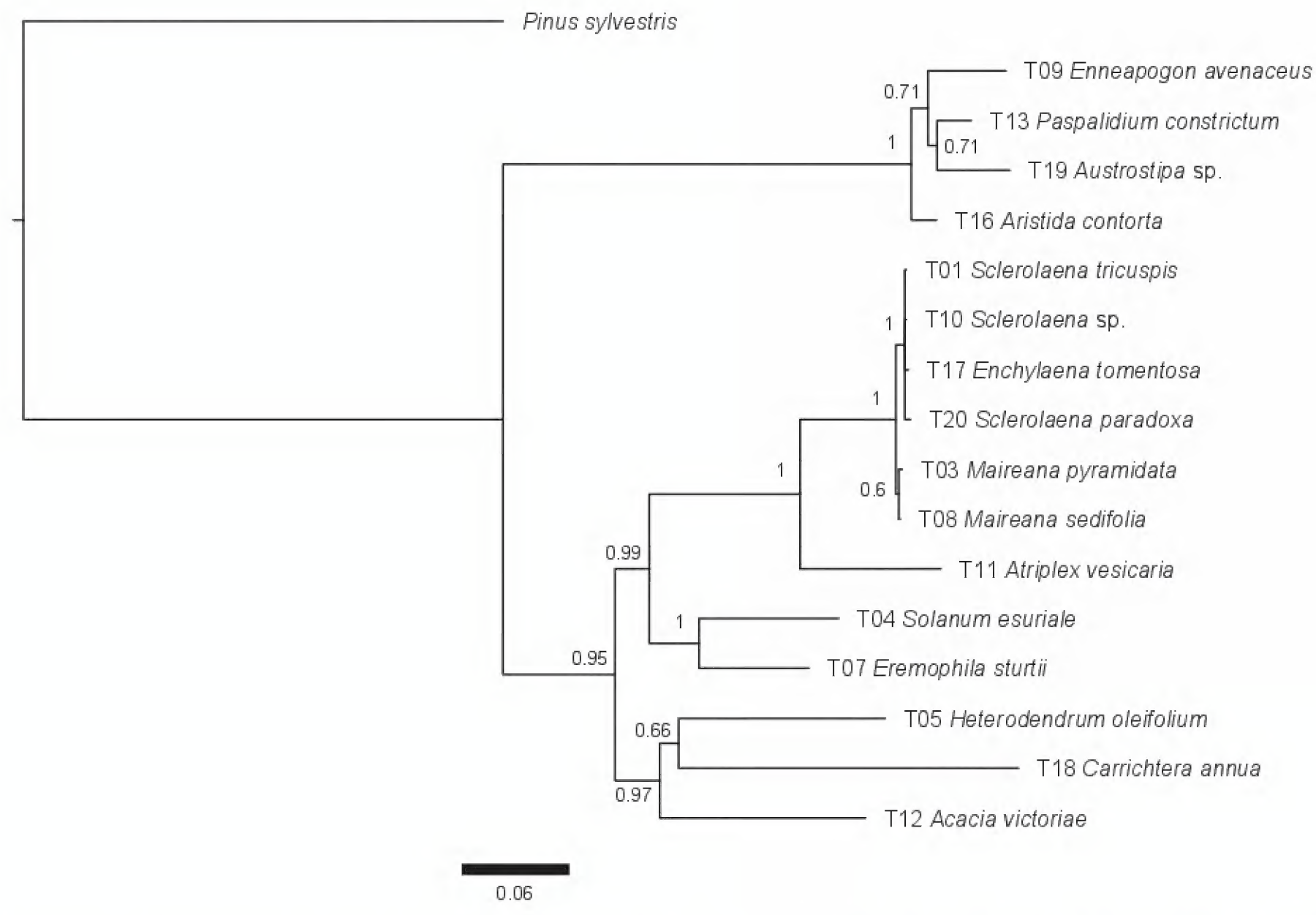

Figure 2. Bayesian tree of *matK* gene sequence data. Numbers at nodes are posterior probability values. Scale bar indicates substitutions per site. The tree was rooted with a single *Pinus sylvestris* sequence downloaded from GenBank (accession number AB097781.1).

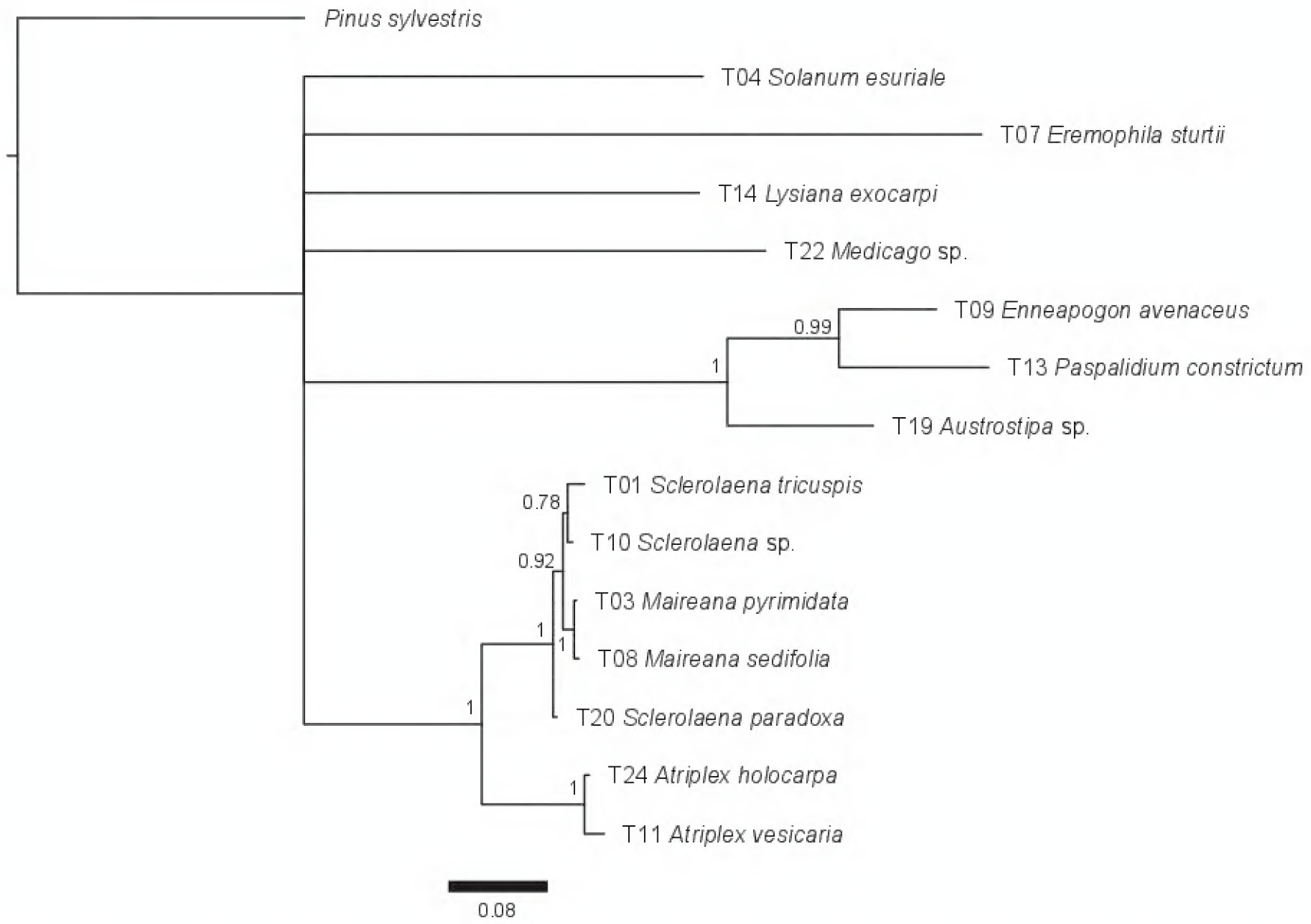

Figure 3. Bayesian tree of *ITS2* sequence data. Numbers at nodes are posterior probability values. Scale bar indicates substitutions per site. The tree was rooted with a single *Pinus sylvestris* sequence downloaded from GenBank (accession number KX167560.1).

Table 5. Family level taxa (GenBank data) ('*' indicates that the column contains no taxa).

Merino Only	Dorper Only	Goat Only	Dorper and Merino	Merino and Goat*	Dorper and Goat	Dorper, Merino and Goat
Casuarinaceae	Caryophyllaceae	Haloragaceae	Lamiaceae		Geraniaceae	Aizoaceae
Elatinaceae	Euphorbiaceae				Marsileaceae	Amaranthaceae
Solanaceae	Nyctaginaceae					Asteraceae
	Plumbaginaceae					Brassicaceae
	Zygophyllaceae					Convolvulaceae
						Fabaceae
						Goodeniaceae
						Malvaceae
						Myrtaceae
						Poaceae

BOLD Results

BOLD database searches yielded very similar results to the GenBank results (Suppl. materials 5, 6). At the family level, there were eight families present in the diets of both sheep breeds, one family found in Merinos not Dorpers, and five families found in Dorpers not Merinos. At the species level, there were 12 species in common between the breeds, four species found in Merinos not Dorpers and seven species found in Dorpers not Merinos. At the family level, goats were found again to have lower diversity than either Merinos or Dorpers, with seven families present. One of these families was unique to goats, three were shared only with Dorpers, and seven families were common to all three taxa. At the species level, the goat diet also was less varied than either sheep, having only seven species present. None of these species was unique to goats, one was shared only with Merinos, one was shared only with Dorpers, and five species were common to all three taxa.

Species and Family Accumulation Curves

To estimate the completeness of taxon sampling, taxon accumulation curves at the family and species level were constructed for each taxon separately based on GenBank results. The species and family accumulation curves indicate that more samples are needed to estimate the total

diversity of Dorper (Suppl. material 7), Merino (Suppl. material 8) and goat (Suppl. material 9) diets, except in two cases—the goat and Merino family-level graphs both asymptote at 13 and 14 families respectively. This indicates that while our sampling of Merinos and goats was probably sufficient to estimate family-level diversity, family-level diversity for Dorpers and species level diversity in all taxa were likely underestimated.

Diet diversity in individual sheep

Dorper and Merino diet data at the species and family levels were found to be normally distributed and with equal variance (Suppl. material 10), and t-tests were therefore selected as appropriate to compare their means. The diversity of the individual Dorper and Merino diets were not significantly different at the species level (t = 0.713, df = 15, p-value = 0.487) nor at the family level (t = 0.579, df = 15, p-value = 0.571).

Among the breeds, the ratio of the total number of contigs to the number of contigs that could be identified to species in each individual was compared. For the ITS2 data, the Kruskal-Wallis test was used, as the data failed the Shapiro-Wilk test for normality, while the ANOVA was used for the rbcL data as this data passed normality and equality of variance tests (Suppl. material 11). In both cases there was no significant differences among the breeds (ITS2, Kruskal-Wallis test, χ 2 = 0.327, df = 2, p-value = 0.849; rbcL, ANOVA, df = 2, F value = 0.085, p-value = 0.919, residual df = 21).

Proportion of Reads in Diet (at family level)

Previous studies have used the proportion of reads from different families recovered in faecal metabarcoding as a proxy for the quantity of different foods in the subjects' diets (Kartzinel et al. 2015, Lopes et al. 2015, Soininen et al. 2015). This analysis was based on the ITS2 GenBank data, as it exhibited the highest diversity.

Our results are highly variable among individuals, but Malvaceae appears to be the dominant family in all three taxa, accounting for between 11% and 34% of the reads (width of standard error, Fig. 4). Amaranthaceae was the next most prevalent among sheep, accounting for between 12% and 28% of the reads, but only 4-8.5% in goats. Standard error was high among samples, indicating the high level of variability in the diet among individuals. ANOSIMs were non-significant using both proportional (R = -0.026, p = 0.632) and presence/absence data (R = -0.061, p = 0.852). The R values close to zero indicate that the most similar samples were found both within-groups and among-groups.

Supplementary analysis at minimum read depth of 3

Decreasing the minimum read depth to 3 increased the number of taxa recovered from the diets of all three study organisms. In the GenBank analysis at a minimum read depth of 3, the sheep have 18 species in common, 12 unique to Merinos and 15 unique to Dorpers; in the BOLD analysis at a read depth of 3, the sheep have 20 species in common, 6 unique to Merinos and 15 unique to Dorpers (Suppl. materials 12, 13, 14, 15). These analyses were similar to the 10 minimum read depth analysis, in that Dorpers exhibited the greatest diet diversity, followed by Merinos and then goats.

Discussion

Both species and family level results show that, when considered at the level of the breed, in aggregate the diet of Dorpers is more varied than that of the Merinos, with family and species level diversity being 29% and 24% higher in Dorpers respectively, based on GenBank results. However, t-test results show that, at the individual sheep diet level, Dorpers do not have a more varied diet than Merinos. This result means that, on average, while an individual Dorper does not have more plant taxa in its diet than an individual Merino, the diet of Dorpers varies more from individual to individual than does that of Merinos. This is consistent with a behavioural model in which Dorpers eat whatever food is easily available without discriminating, while Merinos are more likely to seek out a more restricted set of food plants. This corroborates the results reviewed by Brand (2000) who found that "The Dorper… utilised a larger number of different plant species than Merinos", and that "Dorpers walked less to select food, or a suitable spot to graze". Goats were found to have a less varied diet than either breed of sheep, and the two sheep breeds were found to have a more similar diet to one another than either do to goats, as goats are not eating many of the species that Merinos and Dorpers have in common.

Threatened taxa were found in the diets of both Merinos and Dorpers. Species level matches were found for two taxa in the Dorper diet (*Minuria cunninghamii* and *Sida fibulifera*) and one taxon from the Merino diet (*S. fibulifera*). Genus level matches were also found for five species in the Merino diet, and four species in the Dorper diet, indicating that the diets of these sheep may include more threatened taxa. This highlights the need for barcode sequences from threatened taxa to be uploaded to online databases to facilitate their detection in metabarcoding studies.

The *rbcL*, *matK* and ITS2 sequences could only be successfully amplified in a subset of reference samples. Poor DNA preservation in the plant tissue samples may have been a factor in PCR failure, since trial PCRs of the small amplicons designed for Illumina sequencing were successful for *rbcL* with all 24 samples. However, possible primer- template mismatch cannot be discounted either.

There were several instances of zero or low variation between reference samples from different species, although this is not uncommon in plant DNA barcoding studies (Hollingsworth et al. 2011). This low variation

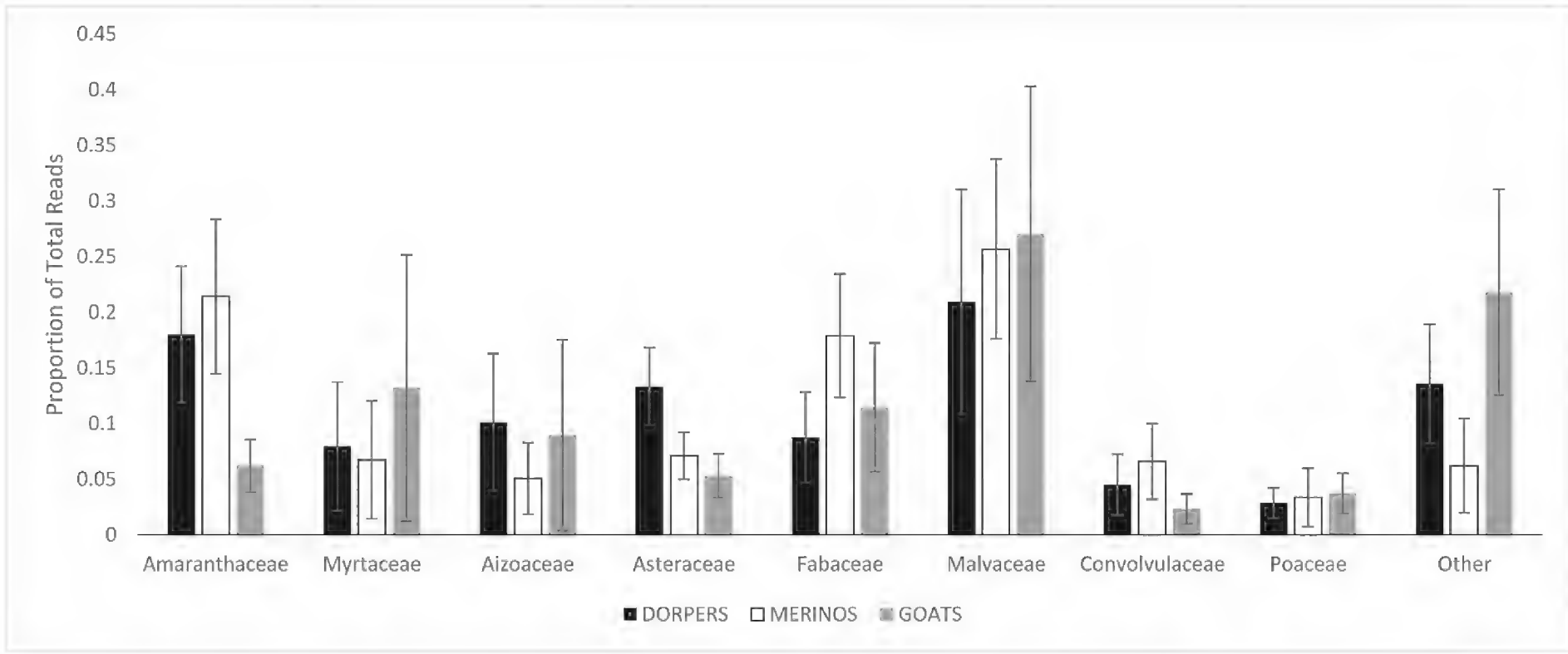

Figure 4. Proportion of reads of each plant family in the ITS2 data from each taxon, based on matches from GenBank. Error bars represent standard error.

may indicate these samples were misidentified morphologically, and are really of the same species. To assess this possibility each sequence was used in a BLAST search of the BOLD and GenBank databases to find their closest matches (Suppl. material 2, Suppl. material 3). Unfortunately, the BOLD database currently contains relatively few plant sequences (95,000 for *rbcL*, 97,000 for ITS2 and 70,000 for *matK*, versus almost 5 million COI barcodes for animals) and none of the genera sampled in this study have yet been sampled comprehensively, which complicates interpretation of the data. Examination of Suppl. material 2 suggests that some identifications may have to be reassessed, e.g. the notoriously difficult *Austrostipa* vs. *Nasella*. However, many of the plant sequences on BOLD are draft sequences that have not yet been published, and also may have been incorrectly identified (e.g. *Chenopodium* ww08577, an *rbcL* match on the BOLD database, is an unpublished sequence which is likely to actually be a species of *Atriplex* based on a number of very close matches to *Atriplex* sequences in BOLD). This identification also accords with our samples T24 and T11, identified using morphology as *Atriplex* species, and for which both *rbcL* and ITS2 data matches *Atriplex* on BOLD. Sequencing of expertly-curated herbarium material is needed to establish firmly both the identity of each species and whether a supplementary marker can reliably distinguish these species.

Although the proportion of reads in the diet metabarcoding dataset is unlikely to match exactly the proportions by volume of plant matter in the diet, there is a relationship between the two (Thomas et al. 2013, Batovska et al. 2017). Using proportions of reads as a proxy for diet content has been used previously in herbivores (Ando et al. 2013). The proportional read data indicates that the three taxa have a similar 'core' diet at the family level, with 55% of the reads being accounted for by Malvaceae, Fabaceae, Amaranthaceae and Aizoaceae in all three taxa. Malvaceae was the most common family in the diet of all three taxa with >20% of the reads in each case. While Dorpers do eat a more diverse diet than Merinos, the additional taxa that they are eating do not form a large part of their diet. The eight families unique to Dorpers together account for only 8-19% of the diet, less than the single most prevalent family, Malvaceae. Brand (2000) reviewed a range of studies that indicated that grasses account for a large part (if highly variable, from 10-75%) of the Merino and Dorper diet, and a greater part of the Merino diet relative to Dorpers. In the present study, grasses accounted for only 2-6% of the reads in the diets of all three taxa, with all three within standard error of one another.

The goat diet was found to be less diverse than the diets of either sheep, and broadly in line with results of previous studies. Goats are known to consume more woody plants and fewer herbaceous plants than sheep (Castro and Fernández Núñez 2016, Bartolomé et al. 1998). In this study, when compared to the sheep, goats were found to consume more plants from the Myrtaceae family of woody plants (although this varied widely among individual goats) and less from the family Amaranthaceae, which are mainly herbaceous.

The range of plants available in the study area at the time the samples were taken may go some way towards explaining these results. The region received more than double the median rainfall during the two years preceding the collection of the dung samples (2010 and 2011). For this reason, plant growth and vegetation diversity were very high, with both Malvaceae and Poaceae being abundant. Merinos may therefore have been able to select only their preferred species as they are also known to consume some of the species found in Dorper-only dung when their preferred species are not present, such as during drought conditions.

It should be noted that, while our proportional reads analysis based on GenBank data showed that Malvaceae was the largest component of the diet of all three taxa, Malvaceae was almost entirely absent from the BOLD

results. This was due to the difference in analytical approach between the two parts of our study. The Malvaceae component of the diets of all three taxa mainly comprised species of the genus *Sida*. The *Sida* sequences on BOLD were shorter than the contigs in our dataset, and as they had less than 100% query coverage, under the method we used in our BOLD analysis those matches were excluded. Matches to *Abutilon*, which was less common, did result in a BOLD match. Care must be taken in interpreting metabarcoding results, as difficulties arise when one is working with reference databases where some but not all of the species that occur in the study area are represented (Zepeda Mendoza et al. 2015). In that case, when one searches the barcode library with a query sequence, the closest matching sequence in the library (even if it is an exact match for a slowly evolving gene like *rbcL*) may not be the same species, or even a congeneric species. Indeed, Wilkinson et al. (2017) caution that the proportion of species- unique *rbcL* + *matK* barcodes is likely to drop even further as sampling increases. To make a judgement call about species identity, one has to know what closely related species are found in the sampling area, and whether those species are represented in your reference library.

Duplicate samples were not highly similar. Based on the GenBank data, the two repeated Dorper samples had 10 species in common, and 18 species possessed by only one sample (Sørensen similarity index = 52.6%). The two repeated Merino samples had 5 species in common, and 11 species possessed by only one sample (Sørensen similarity index = 47.6%). This highlights the value in taking multiple samples from a single individual, as diets are likely to be highly variable. This has been observed previously in another herbivore, the Pacific pocket mouse (Iwanowicz et al. 2016). Only a single sample was taken from each faecal pellet- it might also be useful to compare subsamples from a single pellet to determine its degree of consistency. Regarding the source of the duplicate dung samples, while we are confident that each sample in a vial is of the same sheep, we cannot be absolutely certain of this as there is a small chance that two or more animals in a mob could have dropped their dung at the same spot. Additional sampling, and the use of degenerate primers like the *rbcL*-AM2f primer used in this study, might also help to overcome PCR amplification bias, which can skew the proportion of reads from different taxa in metabarcoding HTS results (Krehenwinkel et al. 2017, Pawluczyk et al. 2015). In our supplementary study increasing the sensitivity of our analysis by lowering the minimum read depth to 3, the number of families and species recovered in Merinos, Dorpers and goats were higher then when using a minimum read depth of 10, but the relative levels of diversity among taxa were maintained.

As sequencing technologies become cheaper and provide more data, the bottleneck in metabarcoding studies becomes whether there exist complete and reliable databases of sequence data from target organisms in the study area. Diet metabarcoding studies on Australian herbivores would be greatly improved by increased taxon sampling and sequencing of Australia's flora. Our results show that much of the diet of the sheep and goats in the area was not covered by our small reference collection, as few barcode IDs matched to our reference database. Incomplete reference collections remain a challenge in carrying out metabarcoding studies (Zepeda Mendoza et al. 2015), and a reliable and complete database of Australian plant taxa would have greatly improved the accuracy of our identifications. In particular, such a database should include those plants listed on threatened species lists to facilitate the identification of herbivores that are potential threats.

Plant barcoding remains challenging for several reasons. Firstly, while the COI gene in animals, being variable, easily amplified and easily aligned, is an excellent standard barcoding region, no single such region exists in plants. There is some diversity of opinion on what plant barcode region is most useful, and most researchers use more than one region at once (Hollingsworth et al. 2016). This inconsistency makes it more difficult to develop and use online sequence databases to identify plant sequences. Including a more variable gene such as *trnL* in this study might have enabled us to recover more matches in the faecal samples, however, more complete *trnL* reference databases are needed to facilitate the identification of recovered sequences.

Genome skimming is a relatively new technique that has the potential to circumvent some of the present difficulties in plant metabarcoding (Dodsworth 2015). First introduced by Straub et al. (2011) in the context of plant systematics, genome skimming provides much more sequence data than barcoding, and might represent a new direction in identifying plant samples from molecular data (Hollingsworth et al. 2016, Wilkinson et al. 2017).

Conclusions

The diet of Dorper sheep was found to be more diverse than the diet of Merino sheep. Although the diet of Dorpers generally was more diverse, individual Dorper sheep do not seem to eat more taxa than Merino sheep, but they do appear to be less selective. The diets of the two sheep breeds were more similar to one another than they were to the goat diet. The proportions of different plant families present in the diets of all three animals show a "core" diet of four plant families, common to all three taxa and accounting for over 55% of the reads. The diversity in the Dorper diet is mainly accounted for by taxa consumed in low quantities. More sampling is required to get a full picture of the diversity of these animals' diets.

Our preparation of a small reference database, and comparison of this database against the online BOLD and GenBank databases, emphasizes that online databases need to be more complete to get a more accurate picture of diversity present. A more complete database would also increase the usefulness of the ITS2 dataset, making it generally more useful than *rbcL* for species-level discrimination.

Finally, Dorpers might be a threat to the vulnerable plant species *Minuria cunninghamii* and *Sida fibulifera*, with Merinos also being a potential threat to *S. fibulifera*. The diets of both sheep breeds also contained taxa identified only to genus, which might potentially include threatened taxa. This highlights that Dorpers could potentially have as much of an impact as Merinos on threatened species.

Acknowledgements

This project is jointly funded by the Lower Murray-Darling Catchment Management Authority (now Western Local Land Services) and the NSW Department of Primary Industries. We are grateful to the many landholders who generously allowed us to undertake the field work in their properties.

References

Alemseged Y, Hacker R (2014) Introduction of Dorper sheep into Australian rangelands: implications for production and natural resource management. The Rangeland Journal 36(1): 85–90. https://doi.org/10.1071/rj13034

Ando H, Setsuko S, Horikoshi K, Suzuki H, Umehara S, Inoue-Murayama M, Isagi Y (2013) Diet analysis by next-generation sequencing indicates the frequent consumption of introduced plants by the critically endangered red-headed wood pigeon (*Columba janthina nitens*) in oceanic island habitats. Ecology and Evolution 3(12): 4057–4069. https://doi.org/10.1002/ece3.773

Atlas of Living Australia (2017) Atlas of Living Australia. http://www.ala.org.au [Accessed on: 2017-8-03]

Bartolomé J, Franch J, Plaixats J, Seligman NG (1998) Diet Selection by Sheep and Goats on Mediterranean Heath-Woodland Range. Journal of Range Management 51(4): 383–391. https://doi.org/10.2307/4003322

Batovska J, Lynch SE, Cogan NOI, Brown K, Darbro JM, Kho EA, Blacket MJ (2017) Effective mosquito and arbovirus surveillance using metabarcoding. Molecular Ecology Resources 18(1)[2018]: 32–40. https://doi.org/10.1111/1755-0998.12682

Benson D, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman D, Ostell J, Sayers E (2012) GenBank. Nucleic Acids Research 41: 36–42. https://doi.org/10.1093/nar/gks1195

Brand TS (2000) Grazing behaviour and diet selection by Dorper sheep. Small Ruminant Research 36(2): 147–158. https://doi.org/10.1016/s0921-4488(99)00158-3

Castro M, Fernández Núñez E (2016) Seasonal grazing of goats and sheep on Mediterranean mountain rangelands of northeast Portugal. Livestock Research for Rural Development 28(5): #91. http://www.lrrd.org/lrrd28/5/cast28091.html

CBOL Plant Working Group, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, der Bank Mv, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim K-, Kress WJ, Schneider H, AlphenStahl Jv, Barrett SC, den Berg Cv, Bogarin D, Burgess KS, Cameron KM, Carine M, Chacon J, Clark A, Clarkson JJ, Conrad F, Devey DS, Ford CS, Hedderson TA, Hollingsworth ML, Husband BC, Kelly LJ, Kesanakurti PR, Kim JS, Kim YD, Lahaye R, Lee HL, Long DG, Madrinan S, Maurin O, Meusnier I, Newmaster SG, Park CW, Percy DM, Petersen G, Richardson JE, Salazar GA, Savolainen V, Seberg O, Wilkinson MJ, Yi DK, Little DP (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences 106(31): 12794–12797. https://doi.org/10.1073/pnas.0905845106

Chase M, Soltis D, Olmstead R, Morgan D, Les D, Mishler B, Duvall M, Price R, Hills H, Qiu Y, Kron K, Rettig J, Conti E, Palmer J, Manhart J, Sytsma K, Michaels H, Kress WJ, Karol K, Clark WD, Hedren M, Gaut B, Jansen R, Kim K, Wimpee C, Smith J, Furnier G, Strauss S, Xiang Q, Plunkett G, Soltis P, Swensen S, Williams S, Gadek P, Quinn C, Eguiarte L, Golenberg E, Learn G, Graham S, Barrett SH, Dayanandan S, Albert V (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL* . Annals of the Missouri Botanical Garden 80(3): 528–580. https://doi.org/10.2307/2399846

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE 5(1): e8613. https://doi.org/10.1371/journal.pone.0008613

Coissac E, Hollingsworth P, Lavergne S, Taberlet P (2016) From barcodes to genomes: extending the concept of DNA barcoding. Molecular Ecology 25(7): 1423–1428. https://doi.org/10.1111/mec.13549

Department of the Environment and Energy Commonwealth of Australia (2018a) Buloke Woodlands of the Riverina and Murray-Darling Depression Bioregions in Community and Species Profile and Threats Database, Department of the Environment and Energy, Canberra. http://www.environment.gov.au/cgi-bin/sprat/public/publicshowcommunity.pl?id=3 [Accessed on: 2018-2-06]

Department of the Environment and Energy Commonwealth of Australia (2018b) EPBC act list of threatened flora, Department of the Environment and Energy, Canberra. http://www.environment.gov.au/cgi-bin/sprat/public/publicthreatenedlist.pl?wanted=flora [Accessed on: 2018-2-06]

Dodsworth S (2015) Genome skimming for next-generation biodiversity analysis. Trends in Plant Science 20(9): 525–527. https://doi.org/10.1016/j.tplants.2015.06.012

Dorper Sheep Society of Australia (2017) Dorpers: Australian history. http://www.dorper.com.au/index.php/dorpers/australian-history [Accessed on: 2017-7-31]

Edgar R, Haas B, Clemente J, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27(16): 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Fazekas A, Burgess K, Kesanakurti P, Graham S, Newmaster S, Husband B, Percy D, Hajibabaei M, Barrett SH (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3 (7): e2802. https://doi.org/10.1371/journal.pone.0002802

Food and Agriculture Organization of the United Nations (2017) FAOSTAT statistics database. http://www.fao.org/faostat/en/#home [Accessed on: 2017-7-31]

Gebremedhin B, Flagstad Ø, Bekele A, Chala D, Bakkestuen V, Boessenkool S, Popp M, Gussarova G, Schrøder-Nielsen A, Nemomissa S, Brochmann C, Stenseth N, Epp L (2016) DNA metabarcoding reveals diet overlap between the endangered Walia ibex and domestic goats - implications for conservation. PLOS ONE 11 (7): e0159133. https://doi.org/10.1371/journal.pone.0159133

Glenn T (2011) Field guide to next-generation DNA sequencers. Molecular Ecology Resources 11(5): 759–769. https://doi.org/10.1111/j.1755-0998.2011.03024.x

Goodwin S, McPherson J, McCombie WR (2016) Coming of age: ten years of next- generation sequencing technologies. Nature Reviews Genetics 17(6): 333–351. https://doi.org/10.1038/nrg.2016.49

Hanner R (2012) Data Standards for BARCODE Records in INSDC (BRIs). BARCODE Data Standards v2.4 http://www.barcodeoflife.org/sites/default/files/DWG%20standards%20revision%202.4%20-%2029%20Feb%2012.pdf

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270(1512): 313–321. https://doi.org/10.1098/rspb.2002.2218

Heckenhauer J, Barfuss MJ, Samuel R (2016) Universal multiplexable *matK* primers for DNA barcoding of angiosperms. Applications in Plant Sciences 4(6): 1500137. https://doi.org/10.3732/apps.1500137

Hibert F, Taberlet P, Chave J, Scotti-Saintagne C, Sabatier D, Richard-Hansen C (2013) Unveiling the diet of elusive rainforest herbivores in Next Generation Sequencing era? The tapir as a case study. PLoS ONE 8(4): e60799. https://doi.org/10.1371/journal.pone.0060799

Hollingsworth P, Graham S, Little D (2011) Choosing and using a plant DNA barcode. PLoS ONE 6(5): e19254. https://doi.org/10.1371/journal.pone.0019254

Hollingsworth P, Li D, der Bank Mv, Twyford A (2016) Telling plant species apart with DNA: from barcodes to genomes. Philosophical Transactions of the Royal Society B: Biological Sciences 371(1702): 20150338. https://doi.org/10.1098/rstb.2015.0338

Höss M, Kohn M, Pääbo S, Knauer F, Schröder W (1992) Excrement analysis by PCR. Nature 359(6392): 199–199. https://doi.org/10.1038/359199a0

Huson D, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H, Tappu R (2016) MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing Data. PLOS Computational Biology 12(6): e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Iwanowicz D, Vandergast A, Cornman R, Adams C, Kohn J, Fisher R, Brehme C (2016) Metabarcoding of fecal samples to determine herbivore diets: a case study of the endangered Pacific pocket mouse. PLoS ONE 11(11): e0165366. https://doi.org/10.1371/journal.pone.0165366

Jarman S, McInnes J, Faux C, Polanowski A, Marthick J, Deagle B, Southwell C, Emmerson L (2013) Adélie Penguin Population Diet Monitoring by Analysis of Food DNA in Scats. PLoS ONE 8(12): e82227. https://doi.org/10.1371/journal.pone.0082227

Kartzinel T, Chen P, Coverdale T, Erickson D, Kress WJ, Kuzmina M, Rubenstein D, Wang W, Pringle R (2015) DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. Proceedings of the National Academy of Sciences 112(26): 8019–8024. https://doi.org/10.1073/pnas.1503283112

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evolutionary Biology 6(29): . https://doi.org/10.1186/1471-2148-6-29

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

King RA, Read DS, Traugott M, Symondson WOC (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. Molecular Ecology 17(4): 947–963. https://doi.org/10.1111/j.1365-294x.2007.03613.x

Krehenwinkel H, Wolf M, Lim JY, Rominger A, Simison W, Gillespie R (2017) Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific Reports 7(1). https://doi.org/10.1038/s41598-017-17333-x

Levy S, Myers R (2016) Advancements in Next-Generation Sequencing. Annual Review of Genomics and Human Genetics 17(1): 95–115. https://doi.org/10.1146/annurev- genom-083115-022413

Lopes CM, Barba MD, Boyer F, Mercier C, Silva Filho PJSd, Heidtmann LM, Galiano D, Kubiak BB, Langone P, Garcias FM, Gielly L, Coissac E, de Freitas TRO, Taberlet P (2015) DNA metabarcoding diet analysis for species with parapatric vs sympatric distribution: a case study on subterranean rodents. Heredity 114(5): 525–536. https://doi.org/10.1038/hdy.2014.109

Mitchell A (2008) DNA barcoding demystified. Australian Journal of Entomology 47(3): 169–173. https://doi.org/10.1111/j.1440-6055.2008.00645.x

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara RB, Simpson G, Solymos P, Stevens MHH, Szoecs E, Wagner H (2017) Vegan: Community Ecology Package. 2.4-3. https://CRAN.R- project.org/package=vegan [Release date: 2017-4-07]

Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas D, Kenfack D, Chuyong G, Cruaud C, Hardy O (2013) How effective are DNA barcodes in the identification of African rainforest trees? PLoS ONE 8(4): e54921. https://doi.org/10.1371/journal.pone.0054921

Pawluczyk M, Weiss J, Links M, Aranguren ME, Wilkinson M, Egea-Cortines M (2015) Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. Analytical and Bioanalytical Chemistry 407(7): 1841–1848. https://doi.org/10.1007/s00216-014-8435-y

Pompanon F, Deagle B, Symondson WC, Brown D, Jarman S, Taberlet P (2011) Who is eating what: diet assessment using next generation sequencing. Molecular Ecology 21(8): 1931–1950. https://doi.org/10.1111/j.1365-294x.2011.05403.x

Ratnasingham S, Hebert PN (2007) BARCODING: bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7(3): 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

R Core Team (2017) R: A language and environment for statistical computing. 3.3.3. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Ronquist F, Teslenko M, der Mark Pv, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard M, Huelsenbeck J (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3): 539–542. https://doi.org/10.1093/sysbio/sys029

Royal Agricultural Society of Tasmania (2016) Australian sheep industry. http://hobartshowground.com.au/wp-content/uploads/2016/09/Sheep-Background- Information.pdf [Accessed on: 2017-7-31]

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Acad-

emy of Sciences 74(12): 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Shehzad W, Riaz T, Nawaz M, Miquel C, Poillot C, Shah S, Pompanon F, Coissac E, Taberlet P (2012) Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (Prionailurus bengalensis) in Pakistan. Molecular Ecology
21 (8): 1951–1965. https://doi.org/10.1111/j.1365-294x.2011.05424.x

Soininen E, Gauthier G, Bilodeau F, Berteaux D, Gielly L, Taberlet P, Gussarova G, Bellemain E, Hassel K, Stenøien H, Epp L, Schrøder-Nielsen A, Brochmann C, Yoccoz N (2015) Highly overlapping winter diet in two sympatric lemming species revealed by DNA metabarcoding. PLOS ONE 10(1): e0115335. https://doi.org/10.1371/journal.pone.0115335

Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebø JH, Ims RA, Yoccoz NG, Taberlet P (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Frontiers in Zoology 6(1): 16. https://doi.org/10.1186/1742-9994-6-16

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2011) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. American Journal of Botany 99(2): 349–364. https://doi.org/10.3732/ajb.1100335

Sugimoto T, Ito TY, Taniguchi T, Lkhagvasuren B, Oyunsuren T, Sakamoto Y, Yamanaka N (2018) Diet of sympatric wild and domestic ungulates in southern Mongolia by DNA barcoding analysis. Journal of Mammalogy. https://doi.org/10.1093/jmammal/gyx182

Symondson WC, Harwood J (2014) Special issue on molecular detection of trophic interactions: Unpicking the tangled bank. Molecular Ecology 23(15): 3601–3604. https://doi.org/10.1111/mec.12831

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next- generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21(8): 2045–2050. https://doi.org/10.1111/j.1365-294x.2012.05470.x

Thomas A, Jarman S, Haman K, Trites A, Deagle B (2013) Improving accuracy of DNA diet estimates using food tissue control materials and an evaluation of proxies for digestion bias. Molecular Ecology 23 (15): 3706–3718. https://doi.org/10.1111/mec.12523

Todd JW, Hansen RM (1973) Plant fragments in the feces of bighorns as indicators of food habits. The Journal of Wildlife Management 37(3): 363–366. https://doi.org/10.2307/3800127

Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F, Gielly L, Cruaud C, Nascetti G, Wincker P, Swenson JE, Taberlet P (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. Molecular Ecology Resources 9(1): 51–60. https://doi.org/10.1111/j.1755-0998.2008.02352.x

Weisberg S, Fox J (2011) An {R} Companion to Applied Regression, Second Edition. Sage, Thousand Oaks, CA. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Wilkinson M, Szabo C, Ford C, Yarom Y, Croxford A, Camp A, Gooding P (2017) Replacing Sanger with Next Generation Sequencing to improve coverage and quality of reference DNA barcodes for plants. Scientific Reports 7: 46040. https://doi.org/10.1038/srep46040

Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang X, Xu H, Zhu Y, Xiao P, Chen S (2010) Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE 5(10): e13102. https://doi.org/10.1371/journal.pone.0013102

Zepeda Mendoza ML, Sicheritz-Pontén T, Gilbert MTP (2015) Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. Briefings in Bioinformatics 16(5): 745–758. https://doi.org/10.1093/bib/bbv001

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. Journal of Computational Biology 7: 203–214. https://doi.org/10.1089/10665270050081478

Supplementary material 1

Sample sites map

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Geographical

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl1

Supplementary material 2

DNA Barcode Results (BOLD) for Plant Reference Samples

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Data table of DNA barcoding matches

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl2

Supplementary material 3

DNA barcode results (GenBank) for plant reference samples

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Dietary

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl3

Supplementary material 4

Number of Reads per Sample

Authors: Timothy Lee

Data type: Statistical data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl4

Supplementary material 5

Family level taxa (BOLD Data)

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Dietary

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl5

Supplementary material 6

Species level taxa (BOLD Data)

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Dietary

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl6

Supplementary material 7

Taxon accumulation curves (Dorpers)

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Taxon accumulation curves

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl7

Supplementary material 8

Taxon accumulation curves (Merinos)

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Taxon accumulation curves

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl8

Supplementary material 9

Taxon accumulation curves (goats)

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Taxon accumulation curves

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl9

Supplementary material 10

Tests for normality and equality of variance

Authors: Timothy Lee

Data type: Statistical tests

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl10

Supplementary material 11

Statistical tests for number of ITS2 and rbcL contigs identified to species as a proportion of all contigs

Authors: Timothy Lee

Data type: Statistical analysis

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl11

Supplementary material 12

Family level taxa (GenBank data), at 3 minimum read depth

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Dietary

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl12

Supplementary material 13

Species level taxa (GenBank Data), at 3 minimum read depth

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Dietary

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl13

Supplementary material 14

Family level taxa (BOLD Data), at 3 minimum read depth

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Dietary

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl14

Supplementary material 15

Species level taxa (BOLD Data), at 3 minimum read depth

Authors: Timothy Lee, Yohannes Alemseged, Andrew Mitchell

Data type: Dietary

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.2.22467.suppl15